Evaluation of low-complexity supervised and unsupervised NILM methods and pre-processing for detection of multistate white goods

Mohammad Khazaei
Electronic & Electrical Engineering
University of Strathclyde
Glasgow, UK
mohammad.khazaei@strath.ac.uk

Lina Stankovic
Electronic & Electrical Engineering
University of Strathclyde
Glasgow, UK
lina.stankovic@strath.ac.uk

Vladimir Stankovic
Electronic & Electrical Engineering
University of Strathclyde
Glasgow, UK
vladimir.stankovic@strath.ac.uk

ABSTRACT

According to recent studies by the BBC and the Scottish Fire and Rescue Service, malfunctioning appliances, especially white goods, were responsible for almost 12,000 fires in Great Britain in just over 3 years, and almost everyday in 2019. The top three "offenders" are washing machines, tumble dryers and dishwashers, hence we will focus on these, generally challenging to disaggregate, appliances in this paper. The first step towards remotely assessing safety in the house, e.g., due to appliances not being switched off or appliance malfunction, is by detecting appliance state and consumption from the NILM result generated from smart meter data. While supervised NILM methods are expected to perform best on the house they were trained on, this is not necessarily the case with transfer learning on unseen houses; unsupervised NILM may be a better option. However, unsupervised methods in general tend to be affected by the noise in the form of unknown appliances, varying power levels and signatures. We evaluate the robustness of three well-performing (based on prior studies) low-complexity NILM algorithms in order to determine appliance state and consumption: Decision Tree and KNN (supervised) and DBSCAN (unsupervised), as well as different algorithms for preprocessing to mitigate the effect of noisy data. These are tested on two datasets with different levels of noise, namely REFIT and REDD datasets, resampled to 1 min resolution.

CCS CONCEPTS

•Computing methodologies~Machine learning~Learning paradigms~Supervised learning~Supervised learning by classification

KEYWORDS

NILM, Load Disaggregation, Event Detection and Classification

NILM'20, November 18, 2020, Virtual Event, Japan

ACM ISBN 978-1-4503-8191-8/20/11…$15.00
https://doi.org/10.1145/3427771.3427850

ACM Reference format:

Mohammad Khazaei, Lina Stankovic and Vladimir Stankovic. 2020. Evaluation of low-complexity supervised and unsupervised NILM methods and pre-processing for detection of multistate white goods. In *The 5th International Workshop on Non-Intrusive Load Monitoring (NILM' 20). November 18, 2020, Virtual Event, Japan. ACM, New York, NY, USA, 5 pages.* https://doi.org/10.1145/3427771.3427850

1 INTRODUCTION

Load disaggregation via non-intrusive load monitoring (NILM) offers a non-intrusive, purely computational, software-based approach to separate aggregate load obtained from a single electricity meter into individual appliance loads and provides a timely opportunity to leverage on investment worldwide on smart metering [1],[2]. Besides the obvious NILM application of meaningful energy feedback, timely detection of malfunctioning appliances without resorting to submetering and leveraging on NILM is promising [3], [4]. Indeed, the BBC has reported that malfunctioning appliances, especially white goods, were responsible for almost 12,000 fires in Great Britain in just over 3 years [5], and the Scottish Fire and Rescue Service reported 340 fires in 2019 alone, caused by tumble dryers, washing machine, fridge/freezers and dishwashers [6]. As shown in [3], malfunction in appliances is reflected when the NILM signature deviates significantly from the actual energy consumption and the deviation matches an anomaly signature. Furthermore, [3] indicates that appliance-level anomaly detection performs best for best performing NILM algorithms, but this was only verified for relatively complex supervised Combinatorial Optimization (CO) [7] and Factorial Hidden Markov Model (FHMM) [8] based NILM from the NILMTK [9], Latent Bayesian Melding [10], Super-state HMM (SSHMM) [11] and unsupervised Graph Signal Processing (GSP) [12] NILM algorithms and anomalies in electrical heater and freezer operation in the REFIT dataset [13].

Motivated by the potential of NILM to identify malfunctioning appliances, as demonstrated in [3] and [4], we propose the following requirements for suitable NILM algorithms: (i) near real-time disaggregation to quickly identify an appliance (at fault), (ii) low complexity so it could potentially be run on a smaller device within the building where the appliance is located instead of in-the-cloud, (iii) for scalability purposes, they should be able to work on a range of buildings, where labelled information is not available for training. Leveraging on NILM algorithms that have shown good

disaggregation performance, we select one method in each of the following categories: supervised (which works best on the house they are trained on but are not always transferable to unseen houses [14]) and unsupervised (suboptimal performance compared to supervised methods but robust to a wide range of datasets where no training information is available). Furthermore, NILM algorithms have been shown to have better performance with some pre- and/or post-processing of the meter data and NILM output, respectively [15]. This paper is organized as follows. Section 2 briefly reviews the NILM algorithms we evaluate here, including the proposed 2-step DT algorithm. Section 3 describes the experimental setup followed by our classification and disaggregation results and conclude in Section 4.

2 NILM ALGORITHMS – BRIEF OVERVIEW

As motivated in Section 1, in this paper we focus on low-complexity algorithms which can be used widely, i.e., with smart meter readings that store only active power measurements, sampled at 1-60 seconds. To this effect, we have narrowed our selection of algorithms that have been shown to work well under the above constraints. These are Decision Trees (DT) and K-Nearest Neighbor (K-NN) for supervised NILM and Density-Based Spatial Clustering of applications with noise (DBSCAN) for unsupervised NILM. We also evaluate the effect of pre-processing and post-processing to improve NILM performance.

2.1 Pre-Processing

Median filtering is common step in pre-processing the raw aggregate power measurements to remove outliers. Length of the median filter windows must be carefully chosen according to the signal attributes, such as granularity, to ensure that relevant events are not lost. For example, in our case, we found heuristically that window sizes of 5 and 3 minutes provided the best results for REFIT [13] and REDD [16] dataset houses, respectively. Then, bilateral graph filtering (GBF) [15] is applied to ensure piecewise smoothness of the power signal. At the end, edge sharpening is used to merge unclear consecutive edges. Edge sharpening is used to merge the consecutive rising edges or the consecutive falling edges caused by state changes lasting more than one sample in the time-series power signal.

2.2 Post-Processing

The multi-state appliances under consideration in this paper have similar operating power levels, but differ in their duty cycles. Using expert knowledge of duty cycle, we therefore group rising and closing ΔP that fit within the duration of the maximum duty cycle.

2.3 Decision Tree (DT)

DT-based NILM is a low-complexity supervised approach, that can be trained using a very small labelled data set. In [15][17], only the difference in two consecutive active power measurements, ΔP, is used as a feature for training. To improve performance, in this paper, we also use active power (P) as an additional training feature.

2.4 Density-Based Spatial Clustering (DBSCAN)

DBSCAN requires only two parameters: ε and the minimum number of points required to form a dense region (minPts). [18] shows that DBSCAN is a viable approach for disaggregating two fridges, where about 81% classification accuracy was obtained with the Eco dataset, downsampled to 1 minute. We only consider ΔP as a feature for DBSCAN.

2.5 K-Nearest Neighbor (KNN)

KNN is a supervised method that requires a labelled trained dataset. K is set based on the validation set, comprising 60% of labelled data. The input consists of the K closest training examples in the feature space. KNN's potential for disaggregating dishwasher and clothes dryer on the AMPds2 dataset were demonstrated in [19], where it was shown that KNN has better overall classification accuracy if we consider both active and reactive power (95%) than only active power (73%). In the absence of reactive power, which is rarely available using commercial smart meters, we use P and ΔP instead as features.

2.6 Disaggregation in two stages

One of the obstacles to detect each appliance is the fact that many appliances have similar consumption power, that is, the features are not discriminative enough. To mitigate this issue, we grouped appliances with the similar ΔP as one category (in this paper, dishwasher and washing machine) to be disaggregated by the algorithms, and then we perform an additional disaggregation step on this subgroup down to individual appliances and then do post-processing as described in Section 2.2.

3 RESULTS AND DISCUSSION

We evaluate the algorithms discussed in Section 2 using active power readings from two open-access datasets, downsampled to 1min resolution: (1) House 1 of the REDD dataset [16] (2) Houses 2 and 3 of the REFIT dataset [14], focusing on dishwasher (DW: Tables 1-6), washer-dryer (WD: Tables 7-9), washing machine (Tables 10-12) and tumble dryer (TD: Tables 13-15).

Table 1: DW performance for REDD House 1 with DT

	PR	RE	F-Score	Acc
No pre-processing	0.65	**0.84**	**0.73**	0.63
Median filtering	0.59	0.83	0.69	0.61
Edge sharpening	**0.67**	0.72	0.69	0.63
Median filtering + edge sharpening	**0.67**	0.63	0.65	0.61
Median filtering + GBF + Edge sharpening	0.64	0.66	0.64	0.60
Benchmark of pre-processing with DT [15]			0.57	0.58
Benchmark of pre-processing [15] with SGSP			0.63	**0.72**

For all results presented, testing was carried out over a full month for both REDD (18/04/2011 - 30/05/2011) and REFIT (01/10/2014 -31/10/2014) houses. The classification and disaggregation evaluation metrics used are F-Score, including Precision (PR) and Recall (RE) and Accuracy (Acc), respectively, as [12]. In each table, we highlight (in bold) the best results. We tested various pre-processing methods applied prior to NILM. All NILM algorithms use post-processing as explained in Section 2.2.

Table 2: DW performance for REDD House 1 with KNN

	PR	RE	F-Score	Acc
No pre-processing	0.65	**0.83**	**0.73**	**0.63**
Median filtering	0.65	0.81	0.72	**0.63**
Edge sharpening	**0.70**	0.61	0.66	0.62
Median filtering + edge sharpening	0.67	0.65	0.66	0.62
Median filtering + GBF + Edge sharpening	0.48	0.72	0.57	0.50

Table 3: DW performance for REDD House 1 with DBSCAN

	PR	RE	F-Score	Acc
No pre-processing	0.46	**0.65**	**0.54**	0.52
Median filtering	0.47	0.11	0.18	0.51
Edge sharpening	**0.59**	0.14	0.23	0.52
Median filtering + edge sharpening	0.58	0.23	0.33	**0.54**
Median filtering + GBF + Edge sharpening	0.47	**0.65**	0.34	0.51

Table 4: DW performance for REFIT House 2 with DT

	PR	RE	F-Score	Acc
No pre-processing	0.63	**0.90**	0.74	0.67
Median filtering	**0.73**	0.79	0.76	0.73
Edge sharpening	0.53	0.58	0.56	0.53
Median filtering + edge sharpening	0.63	0.65	0.64	0.62
Median filtering + GBF + Edge sharpening	0.53	0.72	0.61	0.54
Median filtering + 2-step DT	0.71	0.82	**0.77**	**0.88**
Benchmark of pre-processing with DT [15]			0.73	0.61
Benchmark of pre-processing with SGSP [15]			0.73	0.67

Tables 1 and 4 show both classification and disaggregation performance improvement over [15] due to inclusion of P as a feature over DT for dishwasher and washing machine. Additional performance gain for both classification and disaggregation accuracy is obtained via the proposed 2-step DT. Furthermore, Tables 1, 2 and 4 show improvement in classification performance over the best performing supervised GSP algorithm with pre-processing in [15] for DW.

Table 5: DW performance for REFIT House 2 with KNN

	PR	RE	F-Score	Acc
No pre-processing	**0.62**	**0.87**	**0.72**	**0.65**
Median filtering	**0.62**	0.83	0.70	0.64
Edge sharpening	0.44	0.72	0.55	0.42
Median filtering + edge sharpening	0.48	0.77	0.59	0.46
Median filtering + GBF + Edge sharpening	**0.62**	0.72	0.67	0.63

Table 6: DW performance for REFIT House 2 with DBSCAN

	PR	RE	F-Score	Acc
No pre-processing	0.44	0.63	0.52	0.42
Median filtering	0.48	**0.92**	**0.63**	0.45
Edge sharpening	0.34	0.29	0.31	0.37
Median filtering + edge sharpening	**0.51**	0.47	0.49	**0.51**
Median filtering + GBF + Edge sharpening	0.47	0.39	0.43	0.48

Table 7: WD performance for REDD House 1 with DT

	PR	RE	F-Score	Acc
No pre-processing	**0.94**	**0.98**	**0.96**	**0.96**
Median filtering	0.81	0.84	0.83	0.82
Edge sharpening	0.72	0.67	0.69	0.70
Median filtering + edge sharpening	0.80	0.88	0.84	0.83
Median filtering + GBF + Edge sharpening	0.68	0.94	0.79	0.74
Benchmark of DT [2] (no pre-processing)			0.88	

Table 8: WD performance for REDD House 1 with KNN

	PR	RE	F-Score	Acc
No pre-processing	**0.91**	**0.98**	**0.94**	**0.94**
Median filtering	0.23	0.61	0.33	0.22
Edge sharpening	0.23	0.61	0.33	0.22
Median filtering + edge sharpening	0.20	0.51	0.29	0.25
Median filtering + GBF + Edge sharpening	0.47	0.49	0.48	0.47

Table 9: WD performance for REDD House 1 with DBSCAN

	PR	RE	F-Score	Acc
No pre-processing	**0.63**	**0.71**	**0.67**	**0.64**
Median filtering	0.06	0.02	0.03	0.32
Edge sharpening	0.54	0.22	0.31	0.51
Median filtering + edge sharpening	0.44	0.08	0.14	0.49
Median filtering + GBF + Edge sharpening	0.54	0.10	0.16	0.51

Table 10: WM performance for REFIT House 2 with DT

	PR	RE	F-Score	Acc
No pre-processing	0.18	0.44	0.26	0.18
Median filtering	0.27	0.54	0.36	0.07
Edge sharpening	0.08	0.11	0.09	0.02
Median filtering + edge sharpening	0.32	0.39	0.35	0.31
Median filtering + GBF + Edge sharpening	0.29	0.30	0.29	0.18
Median filtering+2-step DT	**0.48**	**0.56**	**0.52**	**0.52**
Benchmark of DT [2] (no pre-processing)			0.36	

Table 11: WM performance for REFIT House 2 with KNN

	PR	RE	F-Score	Acc
No pre-processing	0.15	0.16	0.16	0.20
Median filtering	0.20	**0.35**	0.25	0.07
Edge sharpening	0.09	0.17	0.12	0.11
Median filtering + edge sharpening	0.26	0.40	0.32	0.20
Median filtering + GBF + Edge sharpening	**0.27**	0.33	**0.30**	**0.27**

Table 12: WM performance for REFIT House 2, DBSCAN

	PR	RE	F-Score	Acc
No pre-processing	**0.12**	**0.76**	**0.21**	0.04
Median filtering	0.11	0.56	0.19	0.07
Edge sharpening	0.07	0.42	0.13	0.04
Median filtering + edge sharpening	0.10	0.57	0.18	**0.24**
Median filtering + GBF + Edge sharpening	**0.12**	0.52	0.16	**0.24**

Tables 4 and 10 show an improvement of 16% in classification performance of the washing machine and 15% improvement in disaggregation performance of the dishwasher with the proposed two-step DT disaggregation.

Table 7 also shows improvement in classification accuracy of WD due to inclusion of P as additional feature compared to the benchmark [2], where DT was used without pre-processing.

As expected, supervised DT and KNN perform better than DBSCAN for all considered appliances. The benefit of pre-processing, especially for improving disaggregation accuracy, is observed clearly where performance is poor, as observed for the unsupervised DBSCAN algorithm (Tables 3, 6, 12) and for the challenging washing machine.

Pre-processing is not beneficial for REDD DW (Tables 1-3) and WD (Tables 7-8) at 1-min sampling resolution since the dataset is relatively less noisy than the REFIT dataset. In fact, it is detrimental because it removes some important edges. However, for the noisier (due to additional unknown appliances) REFIT houses and challenging washing machine, median filtering only is sufficient to improve classification accuracy whilst edge sharpening in addition to median filtering, helps improve the disaggregation accuracy, as observed in Tables 10-12.

Tumble dryer results from REFIT House 3 had good recall results with DT and KNN, comparable with other appliances, as we were able to pick out most instances of the appliance running but some post-processing may be needed to reduce false positives. There were no benchmarks for comparison, so we present results in Appendix A for reference.

4 CONCLUSIONS

In this paper we evaluate the performance of DT, K-NN and DBSCAN algorithms in conjunction with pre-processing (median filtering, Graph Bilateral filtering and edge sharpening) for classification and estimating energy consumption of the top three appliances responsible for fires. This helps us assess which of these simple NILM algorithms to consider for the next step of anomaly detection. Our results indicate that pre-processing can improve the disaggregation performance of unsupervised DBSCAN and for appliances which are challenging to disaggregate, e.g., washing machine. DT has the best classification and disaggregation performance for all appliances of interest, comparable to state-of-the-art algorithms, and needing very little training data. Furthermore, we show that the additional inclusion of aggregate power as a feature in addition to the change in power improves the performance of DT compared to previous literature. We also show improvement over the state-of-the-art with the proposed 2-step DT for improving the performance of the washing machine and dishwasher. DT may not be the best choice for transferability on unseen houses and meeting our scalability criteria, and as such further work on transfer learning with DT is needed.

ACKNOWLEDGEMENTS

This work was partly supported by the European Commission under the 'H2020-EU.3.3.1- Reducing energy consumption and carbon footprint by smart and sustainable use' program topic, according to the Grant Agreement No. 767625.

REFERENCES

[1] Mohammad Khazaei, Lina Stankovic, and Vladimir Stankovic 2019. "Trends and challenges in smart metering analytics." In 2019 MTMI International Conference on Emerging Issues in Business, Technology and Applied Sciences, pp. 111-117.

[2] Kanghang He, Lina Stankovic, Jing Liao, and Vladimir Stankovic. 2016, "Non-intrusive load disaggregation using graph signal processing," IEEE Trans. Smart Grids, vol. 9, pp. 1739-1747.

[3] Haroon Rashid, Vladimir Stankovic, Lina Stankovic, and Pushpendra Singh 2019. "Evaluation of non-intrusive load monitoring algorithms for appliance-level anomaly detection." In Proc. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8325-8329. IEEE

[4] Haroon Rashid, Pushpendra Singh, Vladimir Stankovic, and Lina Stankovic 2019. "Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour?" Elsevier Applied Energy 238: 796-805.

[5] https://www.bbc.co.uk/news/uk-33124925

[6] https://www.firescotland.gov.uk/news-campaigns/news/2020/01/electricity-safety-(7).aspx

[7] George William Hart 1992, "Nonintrusive appliance load monitoring," Proceedings of the IEEE, vol. 80, no. 12, pp. 1870–1891.

[8] J Zico Kolter and Tommi S Jaakkola, 2012 "Approximate inference in additive factorial hmms with application to energy disaggregation.," in AISTATS, vol. 22, pp. 1472–1482

[9] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava 2014. "NILMTK: an open source toolkit for non-intrusive load monitoring." In Proceedings of the 5th International Conference on Future Energy Systems, pp. 265-276.

[10] Mingjun Zhong, Nigel Goddard, and Charles Sutton 2015. "Latent Bayesian melding for integrating individual and population models." In Advances in Neural Information Processing Systems, pp. 3618-3626.

[11] Stephen Makonin, Fred Popowich, Ivan V. Bajić, Bob Gill, and Lyn Bartram 2015. "Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring." IEEE Transactions on Smart Grid 7, no. 6: 2575-2585.

[12] Bochao Zhao, Lina Stankovic, and Vladimir Stankovic 2016. "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing." IEEE Access 4: 1784-1799.

[13] David Murray, Lina Stankovic, and Vladimir Stankovic 2017. "An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study." Scientific Data 4, no. 1: 1-12.

[14] David Murray, Lina Stankovic, Vladimir Stankovic, Srdjan Lulic, and Srdjan Sladojevic. 2019. "Transferability of neural network approaches for low-rate energy disaggregation." In Proc. IEEE ICASSP, Int. Conf. Acoustic, Speech, and Sig. Proc., (ICASSP), Brayton, UK.

[15] Bochao Zhao, Kanghang He, Lina Stankovic, and Vladimir Stankovic 2018. "Improving event-based non-intrusive load monitoring using graph signal processing." IEEE Access 6: 53944-53959.

[16] J. Zico Kolter, and Matthew J. Johnson 2011. "REDD: A public data set for energy disaggregation research." In *Workshop on data mining applications in sustainability* (SIGKDD), San Diego, CA, vol. 25, no. Citeseer, pp. 59-62.

[17] Jing Liao, Georgia Elafoudi, Lina Stankovic, and Vladimir Stankovic, 2014. "Non-intrusive appliance load monitoring using low-resolution smart meter data." In Proc. 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 535-540.

[18] Zhixiang Xu, Ningxuan Guo, Yinan Wang, and Gangfeng Yan 2019. "Identifying Fridge Consumption Non-intrusively based on temporal characteristic." In 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2), pp. 2770-2775. IEEE.

[19] Fitra Hidiyanto and Abdul Halim. 2020. KNN Methods with Varied K, Distance and Training Data to Disaggregate NILM with Similar Load Characteristic. In Proceedings of the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering 2020 (APCORISE 2020). Association for Computing Machinery, New York, NY, USA, 93–99. https://doi.org/10.1145/3400934.3400953

APPENDIX A

Classification and disaggregation performance of tumble dryer in REFIT House 3.

Table 13: TD performance for REFIT House 3 with DT

	PR	RE	F-Score	Acc
No pre-processing	**0.34**	**0.78**	**0.47**	0.10
Median filtering	0.32	0.52	0.40	**0.21**
Edge sharpening	0.23	0.30	0.26	0.14
Median filtering + edge sharpening	0.23	0.43	0.30	0.01
Median filtering + GBF + Edge sharpening	0.25	0.44	0.32	0.06

Table 14: TD performance for REFIT House 3 with KNN

	PR	RE	F-Score	Acc
No pre-processing	**0.34**	**0.65**	**0.44**	**0.17**
Median filtering	0.30	0.60	0.40	0.10
Edge sharpening	0.18	0.32	0.23	0.04
Median filtering + edge sharpening	0.29	0.49	0.36	0.13
Median filtering + GBF + Edge sharpening	0.27	0.47	0.34	0.08

Table 15: TD performance for REFIT House 3 with DBSCAN

	PR	RE	F-Score	Acc
No pre-processing	0.16	**0.36**	**0.22**	0.25
Median filtering	0.16	0.01	0.03	0.47
Edge sharpening	0.05	0.01	0.01	**0.48**
Median filtering + edge sharpening	**0.25**	0.03	0.05	0.47
Median filtering + GBF + Edge sharpening	0.12	0.01	0.02	0.47